

Introduction to Pandas

PANDAS



pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language.

Pandas is most widely used **library** for data analysis, and contains functions for accessing, aggregating, joining, and analyzing data. Its data structure, the DataFrame is analogous to SQL tables or Excel worksheets

<https://pandas.pydata.org/>

PANDAS

Learn . Practice . Upskill at Ybi Foundation

1D Series



axis 0 →

shape: (4,)

2D DataFrame

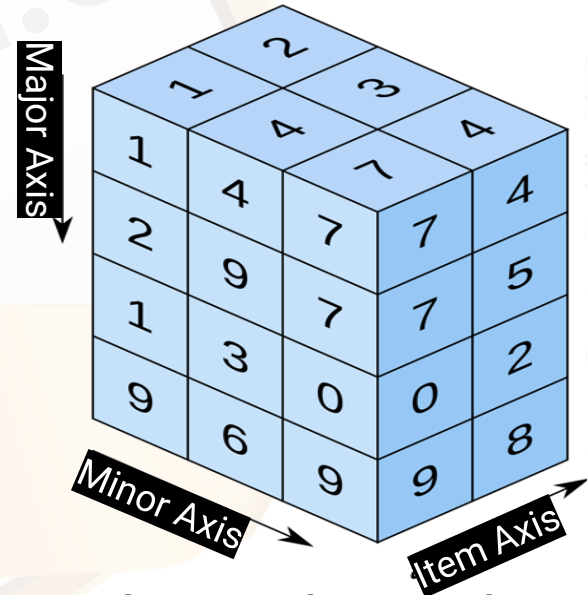


axis 0 ↓

axis 1 →

shape: (2, 3)

3D Panel Data



Major Axis ↓

Minor Axis ↘

Item Axis ↗

shape: (4, 3, 2)

Series and DataFrame

- Pandas series is a one-dimensional array that contains elements of the same data type. Series are mutable means we can modify the elements of a series but its **size is immutable**, i.e. we cannot change the size of the series once declared. It has two main components: data and an index.
- Pandas Dataframe is a 2-dimensional data structure that contains elements of the same data. It is mutable, and its **size is also mutable**, i.e. we can change both data and size of the dataframe data structure. It has labeled axes (rows and columns) and has two different index

SIGNIFICANT FEATURES OF PANDAS LIBRARY?

1. Fast and efficient DataFrame object with default and customized indexing.
2. Read data from different file formats.
3. High-performance merging and joining of data.
4. Handling of missing data.
5. Label-based slicing, indexing, and subsetting of large data sets.
6. Reshaping and pivoting of data sets.
7. Columns from a data structure can be deleted or inserted.
8. Group by data for aggregation and transformations.
9. Time Series functionality.

PANDAS APPLICATIONS

- Read data from different file types
- Manipulation and analyzing data.
- Columns can be inserted and deleted
- Data set merging and joining
- Group by, split, combine data sets
- Flexible reshaping and pivoting of data sets
- Create plots and data visualization
- Easy handling of missing data
- Provides time-series functionality

